ANTICIPATORY TRAFFIC MODELLING AND ROUTE GUIDANCE IN INTELLIGENT VEHICLE/ HIGHWAY SYSTEMS

David E. Kaufman
James Lee
and
Robert L. Smith
Department of Industrial & Operations Engineering
The University of Michigan
Ann Arbor, Michigan 48109-2117

Anticipatory Traffic Modelling and Route Guidance in Intelligent Vehicle/Highway Systems

David E. Kaufman

James Lee

Robert L. Smith

Department of Industrial and Operations Engineering

University of Michigan

Ann Arbor, Michigan 48109

February 27, 1990

Abstract

Traffic modelling has traditionally focused on evaluation of performance of the traffic system in order to determine the aggregate benefits of various possible alterations to the traffic network structure. This paper provides a framework for improving the performance of existing traffic networks by providing real-time route guidance based on traffic conditions changing dynamically in real time and anticipatory projections of short-term future traffic conditions. The value of these projections is that they tend to route vehicles away from an area which is not currently congested but is predicted to be congested by the time the driver would arrive in the area. A number of research issues are identified upon which implementation of the anticipatory model will depend.

1

also reduces discrepancies between routing assignments which minimize individual costs and assignments which minimize total costs.

3 Traffic Assignment Models

The past modelling of traffic has been based primarily on user optimality, also known as Wardrop's first principle, which states that no driver will use a path from his origin to his destination if a shorter path is available to him. If all drivers are in a user-optimal state, the system is said to be in *equilibrium*. The mathematical statement of Wardrop's principle is: all paths used for an O/D at equilibrium have equal cost, and all unused paths have cost greater than or equal to the cost of paths used. This gives rise to a number of closely related mathematical equilibrium models, including nonlinear complementarity, variational inequality, and mathematical programming problems [Friesz, 1985]. The inputs to these models are a network structure, demands for travel between each O/D pair, and expressions for cost of travelling the links of the network as a function of the volume of traffic using each link. The primary outputs produced, when solution techniques are successful, are the volumes of traffic to be found on each link of the network at equilibrium. These models suffer from two poor implicit assumptions: first, that all drivers measure travel costs identically, and second, that all drivers have perfect information on the state of the system.

Furthermore, most of the equilibrium models contain no element of time. The traffic volumes which are input to the cost functions are calculated as if a driver occupies all the links on his route during the entire duration of his trip. Also, no recognition is made of demand for travel varying over time, and link capacity, which is at least implicitly a parameter of the link cost function, is held constant, even though in practice non-recurring congestion occurs every day as a result of traffic incidents, thus reducing capacity.

Traffic assignment models with dynamic elements do exist. A nonlinear mathematical program minimizing total travel cost (*i.e.*, system optimality) can be constructed in a network of many origins and one destination from conservation-of-flow constraints, if for each increment of time t up to some horizon T the following are known: number of drivers departing each origin node at time t, cost functions for each link for each increment of time, and an departure function, determining how many drivers can depart a link in a single time period as a function of the volume on the link [Merchant and Nemhauser, 1978]. By including a decision variable expressing how many cars will be permitted to exit a link, constrained by the departure function, the formulation is extended to make the mathematical program convex and provide explicit information on when traffic can be routed so that it will naturally distribute itself system-optimally (*i.e.*, when the flow permitted to leave each link is equal to the upper bound provided by the departure function at each time increment, so that societal intervention is limited) [Carey, 1987]. Carey also offers generalizations to the

model to include multiple destinations. The weaknesses of this model include the absence of a corresponding user-optimal formulation and the explicit assumption needed for certain results that no link is short enough to be travelled completely in less than a single time increment, which may not apply to network models using certain artificial structures used to describe elements such as turning penalties. Also, the model suffers from end-of-study effects and does not guarantee that each driver reaches his destination.

Some traffic assignment models do not depend on mathematical equilibrium statements and are more heuristic in nature. The chief non-equilibrium methods are the capacity-restraint model and the diversion models. The capacity-restraint model begins by assigning all traffic between an origin/destination (O/D) pair to the shortest path (with respect to travel time) connecting the O/D. Then all travel times are calculated as functions of the resulting traffic volumes, and these travel times are used to adjust the previous traffic assignments iteratively. Among the faults of the method are the absence of any guarantee of convergence of the algorithm and failure to consider limits on roadway capacity.

Diversion models are based on the idea that drivers will not behave according to perfect user optimality in terms of travel time, because drivers do not have perfect information and because preferences vary among drivers and do not depend solely on travel times. The relative volume assigned to a pair of routes depends on some index comparing the routes, such as ratio of travel times, and as the index increasingly favors one route, a greater proportion of traffic is assigned to that route. These models are no longer used as much as they once were; they do not reliably produce assignments which correspond well to measured traffic volumes [Matsoukis, 1986].

Beyond the weaknesses already discussed lies a basic weakness of traffic assignment models. The outputs of these models are the aggregate traffic flows which would occur if a traffic network characterized by uniformity of driver preferences and information, and by fixed capacities, were left to settle into some natural state. But given the constantly changing conditions, the concept of this equilibrium state is of dubious value in real time. The aggregate traffic flows are significant in long-term network planning, but they do not assist a driver in selecting from the possible routes for his O/D; real-time decisions are not improved at all.

To improve quality of travel and reduce congestion, we must better use existing capacity by improving route choice.

4 Dynamic Route Guidance

As is common in traffic modeling, we will assume a known network structure representing the highway system, with a set of nodes I, with N members, and a set A of directed arcs (or links) (i, j) connecting node i to node j. On such a network, if we assume, to begin,

4

that some constant cost (*e.g.*, travel time) c_{ij} for travelling each link (i,j) is known, then the problem of finding the shortest (*i.e.*, least-cost) route from any origin to any destination can easily be solved by a *shortest-path algorithm* applied to a dynamic programming (DP) model of the network. Assume the origin to be labelled node 0 and the destination to be node j, and let f_i be the cost of the least-cost path from node 0 to node i, for any $i \in I$. Then we may use the recursive formulation

$$f_j = \min_{i \neq j} \{f_i + c_{ij}\}$$

Suppose, as seems likely for traffic networks, that $c_{ij} \geq 0$ on all arcs (i,j). Also, let n be the average number of outgoing arcs per node. Then we can solve for the shortest path from node 0 to all nodes $i \in I$ by *Dijkstra's method*, with computational effort no more than $O(Nn \log N)$ [Denardo, 1982], and since n is likely to be a small constant in traffic networks, we may consider the work required to be $O(N \log N)$. For instance, on a machine executing one million instructions per second (MIPS), which would be slow for a personal computer by today's standards, the time required to execute this calculation for a network of 5000 nodes is on the order of 0.1 seconds.

One way to use this model would be to set some average cost c_{ij} on each arc, but this ignores both recurring (demand-based, *e.g.* rushhour) and non-recurring (capacity-based, *e.g.* incidents) congestion. It would be better to have real-time measurement of the state of the traffic network, so that at each point in time the quantities c_{ij} represent the costs of travel on the network in its current state. However, improvement is still possible, because the cost c_{ij} of travelling link (i,j) right now is not entirely relevant if link (i,j) is distant in travel time from the origin node being considered.

Suppose, however, that we possessed quantities $c_{ij}(t)$, which measure the cost of travelling link (i,j) at time t, for $t = 0, \cdots, T$, where the current point in time is $t = 0$, and T is some study horizon, expressed in arbitrary time units (minutes, 2-minute periods, 5-minute periods, etc.). Furthermore, assume that these travel costs consist purely of travel times. Then a new DP can be written, one with *dynamic costs*, by controlling the stage of recursion with the objective function itself. Let f_{ij} be the cost of the least-cost path from node i to node j, departing node i at time $t = 0$, and consider the quantity $c_{kj}(f_{ik})$. Since we are only considering travel times at this point, if we assume that the duration f_{ik} is expressed in the units of t, then $c_{kj}(f_{ik})$ is the time it will take to travel link (k,j) given that we arrive at node k as early as possible, *i.e.*, at the time f_{ik}. Therefore, we may formulate the problem of finding the fastest route from node i to node j by

$$f_{ij} = \min_{k \neq j} \{f_{ik} + c_{kj}(f_{ik})\}$$

and solve for the shortest path by Dijkstra's method *at no additional computational effort* compared to the static shortest-path solution above.
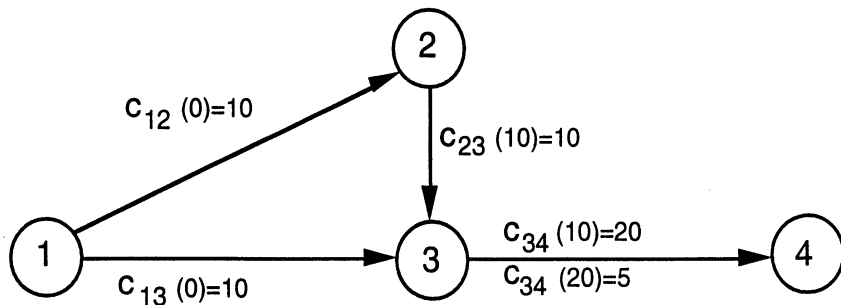
Figure 1: Violation of the principle of optimality

Of course, we can generalize the formulation to include costs other than travel time, by replacing $c_{kj}(f_{ik})$ with $c_{kj}(t_{ik})$, where t_{ik} is the time component of the shortest path with cost f_{ik}. However, for generalized costs, it is possible that the algorithm will find a suboptimal solution, $i.e.$, overlook the least-cost path from node i to node j. To illustrate, consider the network in Figure 1, with the costs as shown. The shortest-path algorithm finds that the shortest path to node 3 is the arc (1,3), with a cost of 10, and at $t = 10$ the cost to go from node 3 to node 4 is 20, for a total cost of 30. However, the path through node 2 has a total cost of only 25. This failure occurs because of a violation of the DP principle of optimality, which states that any route that is part of an optimal route is itself optimal for the nodes it connects. In our example, the violation occurs because a better complete path results from an inferior partial path to node 3. A sufficient condition to prevent the occurrence of suboptimal partial routes is that $c_{ij}(t + s) + s \geq c_{ij}(t)$, which would be true on average for travel times [Prakash, 1989]. To ensure an optimal solution in the absence of this sufficient condition would require T-fold expansion of the state space, with a corresponding increase in computational effort.

An objection to this scheme of routing is that all drivers are receiving the same route, leading to congestion and complete lack of user acceptance of the technology. However, the more frequently the estimates are updated, the more incremental the method becomes; if a number of drivers are routed to the same next link from a certain node within a short period of time, this affects the forecasts, making the link less desirable to the next wave of drivers to reach that node. A further beneficial influence would be the variety of routings resulting from an implementation which allows sufficient customizing of cost functions to individual drivers' preferences. Also, the all-or-nothing routing scheme should be successful in early stages, when market penetration of IVHS technology is low. However, a different strategy may be needed to deal with incidents. An incident can be represented as a discontinuous shift in the capacity of a link, and this may cause a large proportion of the drivers who are very close to the incident location to simultaneously determine identical diversion routes, which would become overloaded; one possible solution would be to have the updates for

a given link calculated very frequently and transmitted frequently to drivers close by but infrequently to drivers far away. It should be noted that incident control is also a primary argument in favor of anticipatory route guidance. When an incident occurs, anticipatory route guidance will immediately start making the region where the incident has occurred less preferable for drivers who might otherwise pass through the area, freeing capacity on the alternate routes around the incident and possibly reducing the duration of the congestion caused by the incident.

When a driver is ready to begin a trip, at $t = 0$, he wants the "best" available route to his destination. If his preferences are accurately represented by the cost functions used to determine the quantities $c_{ij}(t)$, then he can be provided with a minimum-cost route for which the cost on any link is projected based on the point in time when he is expected to travel that link. Toward this end, one can picture an infrastructure which transmits raw traffic data to the vehicles, where a processor customized to the driver's preferences calculates the travel costs and the least-cost path.

This model is a basis for real-time route guidance. Dynamic character is incorporated by the fact that whenever the projections of $c_{ij}(t)$ are altered by unusual congestion behavior or incidents affecting link capacities, a revised minimum-cost path, from the driver's current mid-route location to his original destination, can be computed. Each node is an opportunity for a routing decision, and the anticipatory model allows each driver decision to be the best possible given the information about current and future conditions available at the time of the decision.

5   Anticipatory Traffic Modelling

In the previous section we assumed the existence of projections of $c_{ij}(t)$, the travel costs on the links (i, j) at time t. In this section, we consider how these projections might be developed.

The primary input to these travel costs will be the volumes of traffic projected to be on the various links at time $t = 0, \ldots, T$. These volumes are a major determining factor of travel times, which will often be the main component of travel cost. Let us denote these volumes by $v_{ij}(t)$. Then we can conjecture that $v_{ij}(t)$ will have two main components, one from the segment of the traffic which is in contact (IC) with a Traffic Advisory Center (TAC), and the other from the segment of traffic which is not in contact (NIC). For now, let us assume that the NIC traffic routes itself independently of the routing decisions of the IC traffic; this does not seem too gross, since the NIC traffic has information of an almost entirely static nature.

There are a number of sources of information which might help in projecting volumes $v_{ij}^{NIC}(t)$ due to NIC traffic. One of these would be measurement of current NIC link volumes.

Another would be historical data on the NIC traffic volumes that are typically found trying to use the various links at the time of day for which the projection is being established. Toward these ends, it might be helpful if sensing equipment placed on the links is capable of distinguishing between cars which are accepting information from the TAC and cars which are not. Also of use as an input would be estimation of the demands for travel between an O/D; these can be used as inputs to the existing equilibrium models, which may still have some value in describing flows of traffic with static information.

To investigate projections for IC volumes, we may imagine that once an IC driver selects a route based on dynamic information, his hardware provides that route to the TAC, and that the hardware also transmits the driver's current link and intended route whenever these change. Then we always have the location and route of every IC vehicle in the system. We can, in a sense, track IC vehicles through the future of the traffic system, according to the projected travel times on the links in each vehicle's route. For each t for which we wish to find $v_{ij}(t)$, we then have a picture of where the IC traffic which has already registered on that day should be.

Let us formalize this procedure. Define a *path* k to be any sequence of arcs connecting an O/D and let K be the set of all paths. It will be convenient to refer to links not only by (i,j) but also by ℓ. For any particular path k, let the links on that path be ordered by $\overset{k}{\prec}$ so that the expression $\ell \overset{k}{\prec} (i,j)$ indicates that the links ℓ and (i,j) lie on path k and that on path i, ℓ precedes (i,j). Let h_k be the number of drivers currently routed to path k. To simplify the discussion, let us assume that each IC driver will follow the TAC routing advice perfectly. Then on a given path k, we expect a driver to reach link (i,j) at a time t_{ij} given recursively by

$$t_{ij} = \sum_{\ell \overset{k}{\prec} (i,j)} c_\ell(t_\ell)$$

Then to determine how much previously routed IC traffic should be on link (i,j) at time t, we need to sum the number of drivers on each path who should have gotten to link (i,j) and no farther by t. This may be done by means of an indicator variable:

$$I_{ij}^k(t) = \begin{cases} 1 & \text{if } (i,j) \in k \text{ and} \\ & \quad t_{ij} \leq t < t_{ij} + c_{ij}(t_{ij}) \\ 0 & \text{otherwise} \end{cases}$$

Then we may find the anticipated flows v_{ij}^{IC} by

$$v_{ij}^{IC}(t) = \sum_{k \in K} h_k I_{ij}^k(t)$$

We can then augment this with projections based on historical data of the IC traffic that will enter the traffic system between now and time t (*i.e.*, the IC O/D demands), and we

8

can hope that this historical data will be of a very high quality, since it can be tabulated day by day at the TAC as it is generated by drivers contacting the TAC.

The projections $v_{ij}(t)$ are then simply given by $v_{ij}^{IC}(t) + v_{ij}^{NIC}(t)$. Then, we must have functions which accept the $v_{ij}(t)$ as input and produce as output the projections $c_{ij}(t)$. For a given link (i,j), parameters to the function will include speed limit, number of lanes, quality of pavement, zoning of road (residential, business, controlled access, ...), and real-time information such as time of day and weather conditions. A further input to the function could be the driver generating the request, in order to calculate costs based strictly on that driver's preferences. The accurate formulation of such functions should present a considerable research challenge.

If all this can be achieved, however, then a clear picture of IVHS system operation emerges, with the projections $c_{ij}(t)$ from the current traffic model being used to generate least-cost routes, which in turn are inputs for the next update of the traffic model and the projections $v_{ij}(t)$.

6   Inputs to a Generalized Model

The traffic model we have discussed so far is simplistic in some of its assumptions. This section considers what generalizations are expected to be included in a working model in the framework we have proposed.

The network model of a traffic system as simply a set of nodes and costed directed arcs can be generalized considerably. We can include interaction of traffic at intersections, such as the delays incurred at a signalized intersection with variable signal control strategies. Penalties for making turns can also be included. The model can also be extended to other modes of transportation, i.e., mass transit.

Issues of driver behavior will also be significant. Drivers will not always follow the advice produced by an IVHS system; for instance, they may ignore suggested routes. Also, if in-vehicle hardware is not required to automatically provide the vehicle's current location or if technology does not make this feasible, drivers may begin following a suggested route and then deviate from it without notifying the TAC, reducing the accuracy of the anticipatory projections. This sort of randomness may be of greatest importance when the system is new and drivers are mistrustful. Another behavioral effect concerns the diversion of drivers away from routes slowed by incidents; NIC drivers may observe IC drivers being diverted and follow. Another diversion issue concerns whether diversion is being recommended for reasons of system-optimality or user-optimality. A diversion strategy decreases total travel cost by diverting traffic away from a congested link onto alternate routes. A system-optimal diversion may achieve its benefit by reducing travel cost slightly for the many drivers on the congested link, outweighing a possible large increase in individual travel cost for drivers

who divert; these drivers who are to be diverted may refuse to incur additional personal delays or other costs of the diversion and instead stay on the congested link, anticipating benefits resulting from the diversion of *other* drivers. As a last example, the behavioral issue of varying preferences is very important. The cost of travel includes more than just the time required to complete the trip, but the identity of the other factors and their relative weights will vary widely among drivers. The IVHS system must be able to accommodate a wide variety of preferences to be widely accepted.

In the preceding development, travel times on the links of the network are modelled as simple numerical values, although possibly unknown and requiring forecasting. It may prove more appropriate to model these travel times (and even other forms of travel cost) as random variables, in order to provide drivers with information not only about the expected time required for their trip, but also the variability. A saving in travel time is not always valuable if it results merely in arriving early, but controlling the variability of arrival time may be of great value in avoiding lateness. If we can forecast probability distributions for $c_{ij}(t)$ instead of just numerical estimates, then we can apply *Markov Decision Process* models, which are designed for network models with random costs and which constitute a well-developed subfield of dynamic programming.

Lastly, the model need not be concerned only with providing a route from point A to point B starting immediately. Other services could include suggestions of when to begin the trip, aid in finding parking, or the route for the nearest hospital, for instance.

7 Research Issues

Having discussed the shortcomings of current traffic models and proposed a new modelling environment, we now present a representative sample of research issues which must be addressed in order to achieve a successful implementation of an anticipatory modelling and route guidance system of the nature proposed here.

7.1 Network Aggregation

In what detail must the transportation network be modelled? Must all highways, main arterials, and minor routes be considered or can the system operate with only a subset of all possible routes? An incomplete route may be difficult for a person to follow who is not familiar with the area and does not have a navigational aid.

7.2 Travel Demands

It is easier to measure the traffic volumes on the links of the system at various times than it is to collect data on the O/D demands placed on the network at various times. However, this O/D data is needed to drive the models. There are trip distribution models, such as the gravity model, that are used estimate the O/D demands from the cross-sectional traffic flow data. Research must be done to determine whether these models are sufficiently accurate, how they may be extended, how the O/D demands may be directly measured in an inexpensive manner, and what benefits can be realized by modelling without O/D demand data.

7.3 Driver Behavior

This is an area of great research need. Unknowns include whether a driver will accept a suggested route, when a driver will divert from a suggested route, and the factors which influence these decisions. Even given widespread installation of automatic vehicle location (AVL) technology, we must have some idea of drivers' intentions in order to project future traffic conditions unless models using historical data prove more reliable than they have in the past. Related human factors issues, technology acceptance, methods for improving compliance, and socioeconomic issues affecting driver behavior must also be explored.

7.4 Travel Time/Cost Projections and Assignments

Measuring traffic flow with roadside detection devices can be accomplished, and from this a general idea of travel time on a route segment may be determined. Projecting this traffic flow at some time in the future may use historical values, but these do not consider dynamically changing conditions. Models which project future traffic conditions, given current conditions such as weather, day of week, or whether a large convention or sporting event is taking place, must be developed in order to project travel times and improve vehicle routing.

7.5 Incident Modelling

When an accident occurs, blocking one lane of a three-lane highway, for instance, the model as proposed would react by making a major discontinuous change in the parameter representing the capacity of the affected link. It is not clear whether the routing diversions desirable for this type of event can be handled within the basic framework of the model or whether there is a fundamental theoretical distinction between this and other changes in the anticipatory model which are more incremental in character. Furthermore, it is difficult to predict how long the queueing and congestion caused by the incident will last. This duration is a very

important input to the anticipatory model, because of its effect on the travel time and cost of the affected link over the time horizon being considered. Since incidents are one of the main sources of congestion, the success of IVHS in handling incidents is very important, and the incident models developed will play a major role in determining whether IVHS is a cost-effective tool for improving the entire transportation system.

7.6   Time Horizon

If anticipatory models are required, then the planning horizon, T, must be determined. Each vehicle route interacts with other vehicle routes to a certain degree. Over time and distance while traffic is free flowing this effect is negligible. The effect is not negligible when traffic densities and the chances of incidents are high, $i.e.$, projected variance in travel times are high. Thus, under certain conditions the traffic center may need only project travel costs over a short time period. If a vehicle, however, may be arriving at a location when traffic may be congested, even though the arrival time may be somewhat distant in the future, the proper planning horizon may be much longer. Thus, under what conditions a short horizon versus long planning horizon is necessary is an important research issue.

7.7   Model for Comparison

In analyzing the benefits of any new technology it is necessary to establish a benchmark for comparison. A model must be developed which accurately assesses current performance measures and properly models improvements in these measures due to various implementations of IVHS technology.

7.8   Effect of Market Penetration

When the IVHS system begins to operate, only a small fraction of the driving population will be equipped to use it. In fact, market penetration may even peak at a fairly low level. Therefore, it is necessary to determine the benefits achieved at various market penetration levels. Existing studies indicate that a small change in the traffic volume on a link can have a strong effect on congestion on that link, so considerable benefits may follow from the installation of IVHS technology in a fairly small percentage of cars.

It should be noted that the traffic modelling and analysis component of IVHS research on the issues raised here may be accomplished with few assumptions concerning specific technological developments. The cost and benefit analysis can be done independently of decisions such as how route guidance and other IVHS information is presented to drivers, what type of communication system is used, or how automated vehicle control may operate.

8 Conclusion

A dynamic programming formulation of an anticipatory route guidance system has been presented which shows that a prototype route guidance system may be developed which will also serve as a tool for measuring the value of different IVHS technologies. There are many questions which transportation planners must answer prior to concluding IVHS is a part of the solution to our transportation problems. The research required to properly model traffic within the IVHS environment spans many academic fields and is quite extensive, but will be of value even if many components of IVHS are not implemented. Thus, developing a model which accurately models traffic and driver behavior is required to measure the value of IVHS and should be of highest priority.

9 References

Agnew, W. G. (1988). "Future Personal Ground Transportation", presented at the Symposium on Southern California's Future—An Engineering Challenge.

Carey, M. (1987). "Optimal Time-Varying Flows on Congested Networks," *Operations Research* **35** (1) 58–69.

Denardo, E. (1982). *Dynamic Programming: Models and Applications*. Prentice-Hall (Englewood Cliffs, New Jersey).

Federal Highway Administration (FHWA). (1988). *America's Challenge for Highway Transportation in the 21st Century.*.

Friesz, T. L. (1985). "Transportation Network Equilibrium, Design, and Aggregation: Key Developments and Research Opportunities", *Transportation Research* **19A** (5/6) 413–427.

E. C. Matsoukis (1986). "Road Traffic Assignment - A Review: Part I. Non-Equilibrium Methods," *Transportation Planning and Technology* **11** 69-79.

Merchant, D. K. and Nemhouser, G. L. (1978). "A Model and an Algorithm for the Dynamic Traffic Assignment Problems," *Transportation Science* **12** (3) 183–199.

A. Prakash (1989). Private communication.